Modélisation bayésienne d’une chronologie d’évènements archéologiques
Analyse des chaines de Markov à l’aide du package ’ArchaeoPhases’

Anne Philippe
Marie-Anne Vibet

Laboratoire de Mathémariques Jean Leray
UMR 6629 - Nantes

UNIVERSITÉ DE NANTES

FR n° 2962 du CNRS
Mathématiques des Pays de Loire

Région
PAYS DE LA LOIRE

Introduction

Problématique archéologique et statistique bayésienne

Objectif archéologique

Construire une chronologie d'évènements archéologiques : estimer une succession de dates à partir

- du contexte des évènements (historique, géologique ...)
- de mesures de datation (^{14}C, variation du champ magnétique terrestre, ...)

Pourquoi la statistique bayésienne ?
Prise en compte d'un ensemble d'informations relatives au contexte de l'évènement archéologique

- Contexte historique
  $\rightsquigarrow$ Période d'étude
- Chronologie relative de l'évènement (e.g. stratigraphie)
  $\rightsquigarrow$ Ordre temporel des évènements par rapport aux autres

Exemple de Canimar Abajo (Cuba)

Objectif archéologique : Établir la chronologie des périodes d'utilisation des cimetières de Canimar Abajo (Rocksandic *et al.* 2015)

- de la stratigraphie (superposition des couches)
- des mesures ^{14}C faites sur les os des 12 individus enterrés

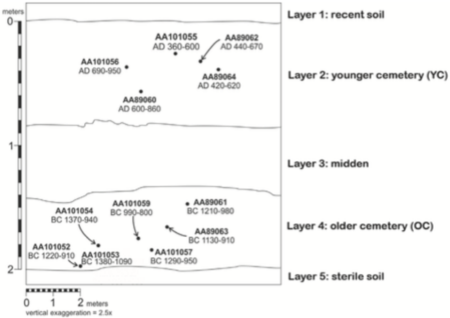


Figure 2 Stratigraphic profile indicating relative positions of samples for AMS ^{14}C dating

Problématique : dater la mort des différents individus enterrés dans les deux cimetières, en déduire les périodes d'utilisation des cimetières et le gap de temps entre les deux cimetières.

Modélisation bayésienne

Qu'est-ce que la statistique bayésienne ?

Information *a priori*

Résumé du contexte de l'évènement

$\rightsquigarrow$ loi de probabilité *a priori*

$$\pi(\theta)$$

Observations

Réalisations d'une loi de probabilité qui dépend du paramètre θ

$$f(\text{data}|\theta)$$

Formule de Bayes

$$\Downarrow$$

Loi *a posteriori*

Mise à jour de la loi du paramètre inconnu θ au vu des observations

$$\pi(\theta|\text{data}) \propto \pi(\theta) \times f(\text{data}|\theta)$$

Échantillonnage par un algorithme de Monte Carlo par chaînes de Markov (MCMC)

Modèle bayésien appliqué à notre exemple

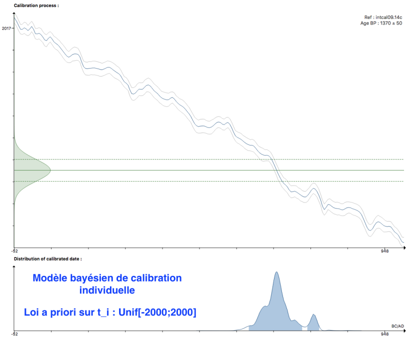
Observations

M_i datation (ici ^{14}C) associée à une erreur de mesure s_i

$$M_i = g(t_i) + \varepsilon_i$$

t_i date de l'évènement archéologique (ici date de mort de l'individu i)
g courbe de calibration

$$\varepsilon_i \sim \mathbf{N}(0, \sigma_g(t_i)^2 + s_i^2)$$



Information *a priori*

$t_1, \ldots, t_5$: dates de mort des individus du cimetière récent (YC),
$t_6, \ldots, t_{12}$: dates de mort des individus du cimetière ancien (OC)
Loi *a priori* des $(t_1, \ldots, t_{12})$: $\{Unif^{12}_{[-2000,2000]} \,|\, t_i < t_j, \forall i = 6, \ldots 12, \forall j = 1, \ldots, 5\}$

=> Loi *a posteriori* : $t_1, \ldots, t_{12} | M_1, \ldots M_{12}$?

Modèle - Language JAGS / Package rjags

JAGS : "boîte noire" permettant de simuler une chaîne de Markov issue de la loi *a posteriori* du paramètre d'intérêt
Paramètres à donner : loi *a priori* du paramètre et loi des observations conditionnelle au paramètre

Programme JAGS

```
"model{
  for( i in 1 : N ) {
  M[i] ~ dnorm(mu[i], tau[i])
  tau[i] <- 1/(s[i]*s[i] + sigma[i]*sigma[i])

  mu[i] <- interp.lin(t[i], Xca, Gca)
  sigma[i] <- interp.lin(t[i], Xca, SDca)
  }
```

Loi des observations

```
  for(i in 1:5){
   t[i]~ dunif(-2000,2000)
  }
  minYC <- min(t[1], t[2], t[3], t[4], t[5])

  for(i in 6:12){
   t[i]~ dunif(-2000,minYC)
  }
```

Loi a priori

```
  }
```

Programme rjags

```
jags <- jags.model("calibration10.jags",
                   data = list('M' = M,'N' = N,'s'=s,'Xca'=Xc
                   n.chains = 2,
                   n.adapt = 100000)
update(jags,100000)

samp = coda.samples(jags,c('t'),100000,thin =10)
```

=> Génération d'un échantillon de taille 20 000 issu de la loi *a posteriori*

Estimation des périodes d’utilisation des cimetières

Analyse des chaines de Markov à l’aide du package ’ArchaeoPhases’

Vérification de la convergence des chaînes de Markov

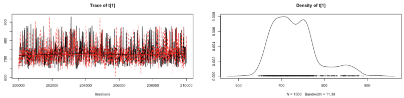
Observation des traces

L'état stationnaire est-il atteint ?

Critère de Gelman-Rubin

Les chaînes ont-elles convergé ?

```
> plot(samp)
```

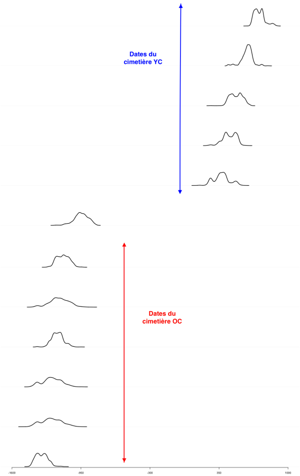



```
> gelman.diag(samp)
Potential scale reduction factors:

      Point est. Upper C.I.
t[1]           1       1.00
t[2]           1       1.01
t[3]           1       1.00
t[4]           1       1.01
t[5]           1       1.01
t[6]           1       1.00
t[7]           1       1.01
t[8]           1       1.00
t[9]           1       1.01
t[10]          1       1.00
t[11]          1       1.00
t[12]          1       1.01

Multivariate psrf

1.01
```

Autres critères : utiliser le package 'coda'

Plus d'information : vignette du package 'ArchaeoPhases'

Chronologie des dates du site de Cuba



Densités marginales des dates t

Package 'ArchaeoPhases' : exploitation de la loi jointe *a posteriori* des dates

$$t_1,...,t_{12}|M_1,...M_{12}$$

=> nouveaux outils pour étudier la dynamique des dates et estimer des périodes de temps.

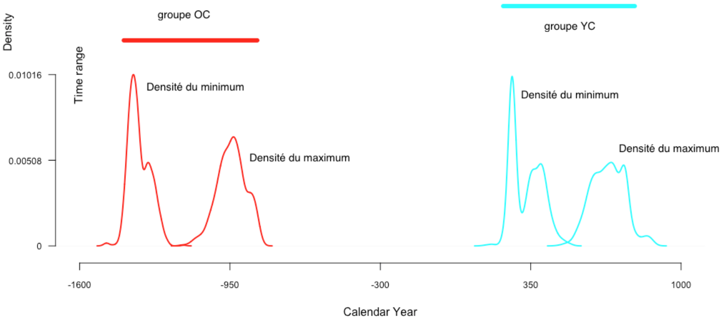
Le début et la fin du groupe de dates sont classiquement caractérisés par les paramètres suivant :

$$\min(t_i, \forall i \in 1,...,5)$$
$$\max(t_i, \forall i \in 1,...,5)$$

Caractérisation des groupes de dates avec 'ArchaeoPhases'

```
MCMC = as.data.frame(rbind(samp[[1]], samp[[2]]))
MinMaxGroup = CreateMinMaxGroup(MCMC, c(1:5), name ="YC", add=NULL, exportFile=NULL)
MinMaxGroup = CreateMinMaxGroup(MCMC, c(6:12), name ="OC", add=MinMaxGroup, exportFile=NULL)
```

Caractéristiques des groupes de dates (1/3)

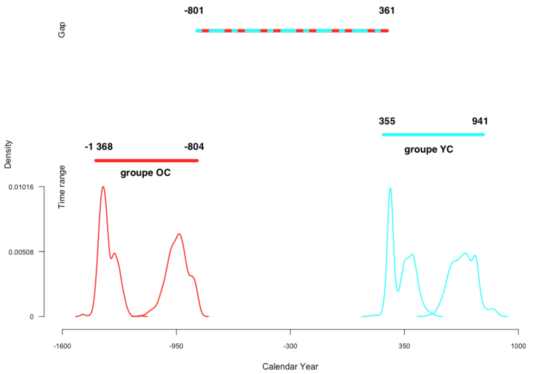



```
MultiPhasePlot(MinMaxGroup, c(3,1), level=0.95)
```

Intervalle de recouvrement d'un groupe de dates à $100 \times (1-\gamma)\%$
Le plus court intervalle $[a,b]$ tel que

$$P(a \leq \text{toutes les dates du groupe} \leq b \,|\, \text{data}) = 1-\gamma$$
$$P(a \leq \min(\text{YC}) < \max(\text{YC}) \leq b \,|\, \text{data}) = 1-\gamma$$

Caractéristiques des groupes de dates (2/3)



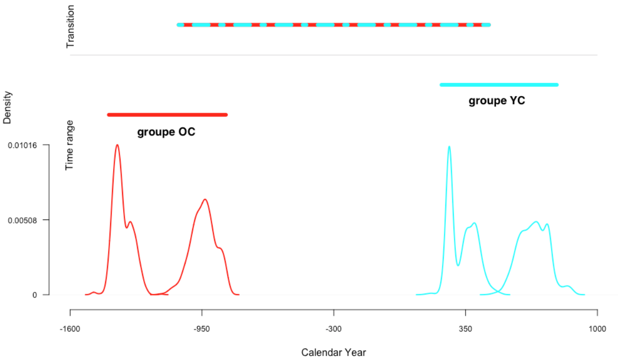

```
MultiSuccessionPlot(MinMaxGroup, c(3,1), level=0.95)
```

Test de l'existence d'un Gap entre deux groupes en succession
S'il existe, alors il s'agit du plus long intervalle tel que

$$P(\text{les dates du groupe OC} \leq a \leq b \leq \text{les dates du groupe YC} \mid \text{data}) = 1 - \gamma$$

$$P(\max(\text{OC}) \leq a \leq b \leq \min(\text{YC}) \mid \text{data}) = 1 - \gamma$$

Caractéristiques des groupes de dates (3/3)




```
MultiSuccessionPlot(MinMaxGroup, c(3,1), level=0.95)
```

Intervalle de transition entre deux groupes en succession
Le plus court intervalle $[a,b]$ tel que

$$P(a \leq \max(\mathrm{OC}) \leq \min(\mathrm{YC}) \leq b \,|\, \mathcal{M}) = 1 - \gamma \tag{3}$$

Conclusions

Résumé

Objectif : Estimer les périodes d'utilisation des deux cimetières superposés de Canimar Abajo (Cuba)

- **Stratigraphie** : 2 couches présentant une activité d'enterrements espacées par une couche constituée de coquillages
- **Observations** : 12 individus datés par radiocarbone

Méthode statistique :
Modélisation bayésienne et simulation à l'aide du package 'rjags'
Estimation des périodes de temps à l'aide du package 'ArchaeoPhases'

Résultats :
Estimation de l'utilisation du cimetière le plus ancien : de -1368 à -804 ans
Estimation de l'utilisation du cimetière le plus récent : de 355 à 941 ans
Estimation du gap : -801 à 361 ans

Bilan des logiciels et packages R

Logiciels de modélisation bayésienne pour construire des chronologies d'évènements archéologiques :

- BCal
- Oxcal
- ChronoModel
- JAGS, Stan

Traitement des chaînes de Markov avec ArchaeoPhases et son application web (démonstration ici)

Packages R pour l'archéologie (liste non-exhaustive) :

- BChron : calibration individuelle, modèle age-profondeur
- ArchaeoPhases : post-traitement des chaines de Markov
- Luminescence : analyse de datations par luminescence
- (A venir) BayLum : analyse bayésienne de datations par luminescence